
Niveau(x) d'enseignement: 0
Centre d'intérêt: 640

Numéro EDUQ : 8670204

La Conception et le développement d'un système informatique qui permet la génération d'un test à partir d'une banque d'items et la passation de ce test de façon interactive / par Denis Rhéaume... [et al.]. -- Sainte-Foy : INRS-Education, 1983.

12 f. ; 28 cm.

"Communication présentée au cinquante-et-unième congrès de l'Association canadienne-française pour l'avancement des sciences à Trois-Rivières, le 25 mai 1983". Code: C-100.

Vedette(s)-secondaire(s): Rhéaume, Denis, 1954- * INRS-Education

Descripteur(s): Elaboration de logiciels * Application pédagogique de l'ordinateur * Enseignement assisté par ordinateur * Elaboration de tests

L'exposé présente un système informatique. Ce système permet d'utiliser l'ordinateur personnel pour: 1. construire une banque d'items; 2. imprimer un test en choisissant au hasard des items de la banque; 3. présenter à l'écran un test composé d'items choisis au hasard permettant ainsi à l'étudiant d'utiliser le terminal pour répondre. Le système peut être utilisé aussi bien à l'école primaire, à l'école secondaire, au cégep ou à l'université. Ses quatre caractéristiques principales sont les suivantes: 1. le système permet un échantillon stratifié d'items; 2. il permet de considérer certaines formes d'items qui appellent une réponse courte et d'aller ainsi au-delà des items à choix de réponses; 3. il comporte un dispositif faisant en sorte que des items jugés incompatibles ne fassent pas partie du même test; 4. il permet à l'étudiant de revenir à un item antérieur pour répondre ou changer sa réponse comme s'il s'agissait d'un test écrit.

Type de document: r

Denis Rhéaume, Normand Dufour, Marie
Makdissi et Yves Bégin.

LA CONCEPTION ET LE DEVELOPPEMENT D'UN
SYSTEME INFORMATIQUE QUI PERMET LA GE-
NERATION D'UN TEST A PARTIR D'UNE ...

INRS-Education, C-100, 1983.

LA CONCEPTION ET LE DÉVELOPPÉMENT D'UN SYSTÈME INFORMATIQUE
QUI PERMET LA GÉNÉRATION D'UN TEST À PARTIR D'UNE BANQUE D'ITEMS
ET LA PASSATION DE CE TEST DE FAÇON INTERACTIVE

par

Denis Rhéaume
Normand Dufour
Marie Makdissi
Yves Bégin

C-100

Communication présentée au cinquante-et-unième
congrès de l'Association canadienne-française
pour l'avancement des sciences
à Trois-Rivières, le 25 mai 1983

INRS-Education
2383 Chemin Ste-Foy
Sainte-Foy (Québec)
G1V 1T1

La conception et le développement d'un système informatique
qui permet la génération d'un test à partir d'une banque d'items
et la passation de ce test de façon interactive

Le but de cet exposé est de vous présenter un système informatique qui permet la génération d'un test à partir d'une banque d'items et la passation de ce test de façon interactive.

Ce système a été conçu et est actuellement développé par une équipe de recherche de l'INRS-Education. La mise au point d'un tel système et son évaluation constitue un des volets de nos travaux ayant trait aux applications pédagogiques de l'ordinateur personnel.

Avant d'aller plus loin, il est important de préciser que les travaux de développement du système présenté aujourd'hui s'étendent sur une période de deux ans ; la première année est consacrée à la production du système alors que la deuxième année prévoit une évaluation du système et de son utilité dans le contexte réel de la classe. Actuellement, nous sommes enclins à croire que le système pourra être utilisé de façon expérimentale durant l'année scolaire 1983-1984.

Le système développé permet d'utiliser l'ordinateur personnel pour :

- (1) construire une banque d'items ;
- (2) imprimer un test en choisissant au hasard des items de la banque;
- (3) présenter à l'écran un test composé d'items choisis au hasard permettant ainsi à l'étudiant d'utiliser le terminal pour répondre.

Notre système peut être utilisé aussi bien à l'école primaire, à l'école secondaire, au cégep ou à l'université.

Le système développé n'est pas limité à une matière en particulier ni à un groupe de matières ; il peut être utilisé lorsque les items que l'on utilise pour la mesure du rendement sont des items à choix de réponses ou des items qui appellent une réponse très courte et une correction objective.

Qui peut être l'utilisateur d'un tel système? Disons d'abord que nous avons voulu produire un système suffisamment simple à utiliser pour pouvoir l'être par des individus qui n'auraient aucune familiarité avec cette technologie nouvelle que représente l'ordinateur personnel. En fait, nous avons accordé beaucoup d'attention à la qualité de la relation entre l'utilisateur et l'ordinateur, et nous continuerons de le faire.

Concrètement, l'utilisateur pourrait très bien être l'enseignant qui désire s'en servir avec ses propres étudiants. Il pourrait aussi s'agir d'un groupe d'enseignants et d'enseignantes désirant mettre en commun leurs efforts pour la production d'une banque d'items. Enfin, il pourrait être utilisé par des individus désirant produire des tests qu'ils mettraient ensuite à la disposition de ceux et celles que cela intéresserait. Notre système est un logiciel pour la production de tests.

QUATRE CARACTÉRISTIQUES IMPORTANTES

Déjà, il existe plusieurs systèmes poursuivant des objectifs semblables aux nôtres et nous en étions d'ailleurs conscients au point de départ. Nous avons néanmoins voulu que le système développé à l'INRS-Éducation comporte certaines caractéristiques que nous jugeons importantes mais que nous ne retrouvions pas ailleurs.

Je mentionnerai donc quatre de ces caractéristiques.

1. Le système permet un échantillon stratifié d'items

En fait, à chaque fois que l'utilisateur introduit un item dans la banque, il doit lui attribuer un numéro de catégorie. Une catégorie représente un groupe d'items, un sous-ensemble de la banque ; chaque catégorie peut correspondre à une section de la matière, à un objectif pédagogique ou à tout autre classement jugé utile.

Au moment de générer un test, l'utilisateur peut déterminer le nombre d'items devant provenir de chaque catégorie, ce nombre pouvant aussi varier d'une catégorie à l'autre. Le test est alors la juxtaposition de plusieurs sous-tests composés d'items choisis au hasard et c'est ce qui nous amène à parler d'échantillon stratifié d'items. Le système permet aussi à l'utilisateur de ne pas tenir compte du classement des items en catégories et d'obtenir un seul échantillon d'items.

L'avantage d'avoir la possibilité de l'échantillon stratifié est grand. Premièrement, il ne doit plus exister de relation parfaite entre l'importance d'une catégorie et le nombre d'items disponibles dans la banque pour cette catégorie. Deuxièmement, l'utilisateur peut

permettre à une catégorie de varier en importance d'une occasion à l'autre sans devoir ajouter ou retrancher des items dans la banque. Troisièmement, il permet de s'assurer d'une meilleure représentativité de l'ensemble des items formant le test.

Notre système d'échantillon stratifié est relativement souple. Le nombre de catégories différentes peut atteindre vingt-cinq. Le nombre d'items pouvant appartenir à une catégorie n'est pas limité, sinon par la taille de la banque elle-même. L'utilisateur peut changer en tout temps la catégorie d'appartenance d'un item.

2. Le système permet de considérer certaines formes d'items qui appellent une réponse courte et d'aller ainsi au-delà des items à choix de réponses

Nous étions conscients au point de départ qu'il était beaucoup plus facile de produire un système ne permettant de considérer que des items pour lesquels l'étudiant ou l'étudiante choisit la lettre ou le chiffre correspondant à sa réponse. Nous avons cependant voulu faire appel à cette possibilité qu'offre l'ordinateur lorsqu'il s'agit de corriger des items qui appellent une réponse courte et non uniquement un choix de réponses.

Le système prévoit donc sept types d'items :

type 1 : choix multiples

type 2 : vrai-faux

type 3 : réponse simple (forme 1)

L'étudiant ou l'étudiante répond en donnant une chaîne de caractères (exemple : un mot, un chiffre). Nous appelons ici une chaîne, toute suite de lettres ou de chiffres juxtaposés. Aucune réponse n'est donc proposée. Une seule chaîne est considérée comme étant la bonne réponse.

type 4 : réponse simple (forme 2)

Encore une fois, l'étudiant ou l'étudiante répond en donnant une seule chaîne de caractères. Cette fois, deux ou trois chaînes peuvent être considérées comme étant la bonne réponse, ce qui permet une correction plus subtile. Nous croyons même qu'une telle possibilité est parfois essentielle. Voici un exemple : si l'on demande à l'étudiant de traduire le mot "automne" en anglais, deux réponses doivent être considérées comme acceptables, "fall" et "autumn".

type 5 : réponse plus longue (forme 1)

Cette fois, la réponse de l'étudiant ou de l'étudiante peut comporter plusieurs chaînes de caractères mais elle ne doit pas dépasser 40 caractères. Il s'agit donc toujours d'items qui appellent une réponse courte.

Aucune réponse n'est proposée. Pour être jugée bonne, la réponse doit contenir une chaîne donnée laquelle est fournie à l'ordinateur comme clé de correction. Voici un exemple d'utilisation : si l'on demande de nommer la rivière la plus importante qui serpente le territoire de la ville de Québec, l'on pourra considérer qu'il y a bonne réponse à partir du moment où l'ordinateur rencontre la chaîne "charles" laquelle doit nécessairement être contenue dans la réponse de l'étudiant qui sait que cette rivière est la rivière Saint-Charles.

type 6 : réponse plus longue (forme 2)

Encore une fois, la réponse de l'étudiant ou de l'étudiante peut comporter plusieurs chaînes de caractères mais ne doit pas dépasser 40 caractères. Cette fois cependant, l'ordinateur peut prendre en considération deux ou trois chaînes comme clé de correction ; la réponse doit contenir l'une ou l'autre des deux ou trois chaînes pour être considérée comme étant bonne.

type 7 : estimé numérique

Pour ce septième et dernier type, encore une fois, aucune réponse n'est proposée à l'étudiant ou à l'étudiante. Sa réponse doit

être un chiffre ; ce chiffre peut être accompagné ou non d'un signe ; il peut comporter une partie décimale ou non. Pour ce type, l'ordinateur considère qu'il y a bonne réponse si la réponse fournie se situe entre une limite inférieure et une limite supérieure lesquelles sont fournies par l'utilisateur comme clé de correction.

C'est surtout durant la phase d'évaluation des travaux de développement que nous verrons les problèmes et les avantages que soulève l'utilisation d'autant de types d'items. Mentionnons par ailleurs que le type de l'item n'a de l'importance que lorsque l'on prévoit que le test sera présenté de façon interactive à l'écran et au terminal. En effet, si le test est imprimé, le problème de la correction par ordinateur ne se pose plus, à moins qu'on tienne à utiliser le lecteur optique. Pour le test imprimé, la prise en compte par le système d'une variété de types d'items est nécessairement avantageuse.

2. Le système comporte un dispositif faisant en sorte que des items jugés incompatibles ne fassent pas partie d'un même test

Notre expérience ayant trait à l'utilisation de banques d'items pour la génération de tests nous a graduellement amenés à voir que dans plusieurs cas, la coapparition de certains items dans un même test engendrait certains problèmes de validité. Mentionnons les deux problèmes les plus courants qui peuvent survenir. Premièrement, la présence d'un item donné dans le test peut contenir un indice important de la réponse à un autre item, ce qui est évidemment habituellement non désirable.

Deuxièmement, il peut arriver que l'on désire placer dans une même banque deux items de forme différente mais touchant trop le même aspect pour que leur coapparition n'apparaisse pas comme clairement injustifiée.

La solution des problèmes évoqués plus haut peut résider évidemment dans le fait de ne jamais mettre dans une même banque des items jugés incompatibles et de ce fait, exclure systématiquement certains items pour cette raison. Nous avons cependant pensé qu'il était à propos d'inclure à notre système un dispositif faisant en sorte que des items de la banque jugés incompatibles ne fassent pas partie d'un même test. Cette façon de faire devrait permettre de conserver dans les banques d'items plusieurs items de qualité lesquels auraient normalement dû être éliminés.

L'introduction d'un tel dispositif, simple en apparence, a engendré plusieurs difficultés au niveau de la programmation, lesquelles ont cependant déjà toutes été surmontées.

Nous verrons mieux lors de la phase d'évaluation l'utilisation qui peut être faite d'un tel dispositif. Mentionnons enfin que ce dispositif est utile non seulement lorsque le test est imprimé mais aussi lorsqu'il est présenté à l'écran.

4. Le système permet à l'étudiant de revenir à un item antérieur pour répondre ou changer sa réponse comme s'il s'agissait d'un test écrit

Cette caractéristique du système ne concerne que le cas où le test est présenté à l'écran. Au point de départ, nous avons le choix entre deux avenues. La première, la plus utilisée par les systèmes

semblables aux nôtres, consistait à fournir à l'étudiant ou à l'étudiante un feed-back immédiat lorsque sa réponse est donnée. La deuxième, celle que nous avons retenue, consiste à permettre de répondre à tous les items du test avant de donner le résultat. Nous avons choisi cette deuxième avenue parce que c'est celle qui de toute évidence s'apparente le plus au test écrit et parce que nous sommes encore enclins à croire que le feed-back à la fin du test permet d'obtenir un résultat peut-être plus valide dans certains cas, étant donné que le feed-back après chaque item équivaut à fournir de l'information durant la passation du test.

À partir du moment où nous retenions la deuxième avenue, il devenait souhaitable d'exiger que l'ordinateur offre au minimum tous les avantages du test écrit et d'une façon particulière, la possibilité pour l'étudiant ou l'étudiante de revenir à un item antérieur pour répondre ou changer sa réponse.

Avec notre système, l'étudiant ou l'étudiante peut identifier ses bonnes réponses et ses erreurs après avoir connu son résultat.

TROIS CARACTÉRISTIQUES SECONDAIRES

Je voudrais maintenant mentionner trois caractéristiques secondaires du système en développement. Je les qualifie de secondaires uniquement parce qu'elles ne concernent pas proprement le domaine touché par notre application, soit la mesure du rendement scolaire. Elles sont toutefois bien importantes dans mon esprit.

Premièrement, le système développé permet d'utiliser pour le texte des items les minuscules et les accents sur les minuscules. Il respecte donc en bonne partie les exigences de la langue française, bien que par exemple, les accents sur les majuscules ne soient pas encore possibles pour nous.

Deuxièmement, le texte des items introduits dans la banque peut être sensiblement plus long que ce qu'offrent plusieurs systèmes comparables au nôtre. Il peut atteindre plus de 400 caractères. Cette possibilité permet de rendre utilisables bon nombre d'items qui devraient être rejetés avec d'autres systèmes ; elle est particulièrement utile pour les items à choix multiples lorsque les réponses proposées à l'étudiant ou à l'étudiante sont longues.

Troisièmement, le système développé comporte un dispositif simple d'utilisation mais relativement puissant permettant de modifier ou de corriger à n'importe quel moment le texte d'un item. Déjà, nous avons l'évidence de la rentabilité d'un tel dispositif, et ce, particulièrement lorsque les items sont longs, ce que permet notre système.

LIMITES DU SYSTÈME

Le système développé n'est pas sans limites, loin de là. La plupart d'entre elles ont surtout trait à l'équipement utilisé. Cette affirmation nécessite que j'indique que l'on prévoit actuellement fonctionner avec un ordinateur personnel Apple II de 48K, un seul lecteur de disquette conventionnel et une imprimante ; ajoutons que l'écran en est un de 40 colonnes.

Actuellement, le nombre maximum d'items pouvant être inclus dans la banque est de 160. Le temps de réponse de l'ordinateur est parfois lent pour certaines opérations. L'utilisation de 40 colonnes pour l'impression du test pose certains problèmes de gaspillage et même d'élégance. La taille d'un test ne doit pas dépasser 50 items, et ce, peu importe leur longueur.

D'une certaine façon, nous considérons que nos principales limites ont trait à l'équipement mais que celles-ci seront rapidement surmontées grâce aux développements technologiques déjà envisagés. Il est par ailleurs intéressant de voir tout ce qui peut être fait avec un équipement de potentiel réduit, et par-dessus tout, il nous semble que des efforts de recherche et de développement considérables doivent être déployés maintenant et qu'il est sage de prévoir l'arrivée des nouveaux outils.

UTILITÉ DU SYSTÈME

Quelques mots maintenant concernant l'utilité du système développé. Disons d'abord que c'est surtout durant la phase d'évaluation du projet que nous serons le mieux en mesure de voir quelles sont les possibilités du système qui apparaissent les plus utiles.

Avant tout, notre système en est un qui permet de construire une banque d'items et de construire un test en choisissant au hasard des items de cette banque. Pour ce faire, l'ordinateur n'est pas requis ; il permet cependant de faire en très peu de temps ce qui serait normalement très long et à peine recommandable. L'utilisation d'une banque

d'items et d'un programme de génération de tests rend possible la production presque automatique de formes parallèles d'un test. Un étudiant ou une étudiante peut avoir la possibilité de passer plus d'une fois un test référant au même contenu mais comportant des items différents.

Là où l'ordinateur est essentiel, c'est pour la présentation du test à l'écran et la correction automatisée. Avec un tel outil, si le contexte de la classe le permet, cela signifie qu'un étudiant ou une étudiante peut se présenter à l'ordinateur au moment qu'il ou elle choisit et subir un test répondant à certaines caractéristiques précises. La disponibilité d'un tel outil rend davantage possible, nous aimerions croire, le développement d'habitudes d'auto-évaluation. L'étudiant ou l'étudiante peut se présenter à plusieurs reprises pour un même test. Il est le seul ou elle est la seule à connaître son résultat et à voir ses erreurs, ce qui peut dans certains cas être psychologiquement très confortable. Disons que même s'il ne permet pas le diagnostic, le système actuellement en développement pourrait s'avérer un outil approprié et apprécié dans un contexte d'évaluation formative et que c'est beaucoup ce que nous entendons vérifier lors de la deuxième partie de nos travaux.